THE SHIFTING TECHNIQUE FOR SOLVING A NONSYMMETRIC ALGEBRAIC RICCATI EQUATION *

CHUN-YUEH CHIANG† AND MATTHEW M. LIN ‡

Abstract.

This paper analyzes a special instance of nonsymmetric algebraic matrix Riccati equations arising from transport theory. Traditional approaches for finding the minimal nonnegative solution of the matrix Riccati equations are based on the fixed point iteration and the speed of the convergence is linear. Relying on simultaneously matrix computation, a structure-preserving doubling algorithm (SDA) with quadratic convergence is designed for improving the speed of convergence. The difficulty is that the double algorithm with quadratic convergence cannot guarantee to work all the time. Our main trust in this work is to show that applied with a suitable shifted technique, the SDA is guaranteed to converge quadratically with no breakdown. Also, we modify the conventional simple iteration algorithm in the critical case to dramatically improve the speed of convergence. Numerical experiments strongly suggest that the total number of computational steps can be significantly reduced via the shifting procedure.

Keywords. nonsymmetric algebraic Riccati equation, transport theory, shifting technique, critical case, structured doubling algorithm, simple iteration method

AMS subject classifications. 15A24, 65F10

1. Introduction. The nonsymmeric algebraic Riccati equation (NARE), encountered in transport theory, is given by

$$XCX - XD - AX + B = 0, \tag{1.1}$$

where A, B, C and $D \in \mathbb{R}^{n \times n}$ are given by

$$A = \Delta - eq^{\top}, \quad B = ee^{\top}, \quad C = qq^{\top}, \quad D = \Gamma - qe^{\top}. \tag{1.2}$$

where

$$
\begin{array}{llll}
e & = & [1, \ldots, 1]^{\top} \in \mathbb{R}^{n}, & \\
q & = & [q_1, \ldots, q_n]^{\top}, & \text{with} \quad q_i = \frac{c_i}{2\omega_i}, \\
\Delta & = & \text{diag}([\delta_1, \ldots, \delta_n]), & \text{with} \quad \delta_i = \frac{1}{c\omega_i(1+\alpha)}, \\
\Gamma & = & \text{diag}([d_1, \ldots, d_n]), & \text{with} \quad d_i = \frac{1}{c\omega_i(1-\alpha)}.
\end{array}
\tag{1.3}
$$

The parameters, used to define the above matrices and vectors, satisfy $0 < c \le 1$, $0 \le \alpha < 1$ and the sequences are $0 < \omega_n < \cdots < \omega_2 < \omega_1 < 1$, $c_i > 0$, $i = 1, 2, \ldots, n$, so that $\sum_{i=1}^{n} c_i = 1$.

For the physical meaning of the NARE (1.1) and its corresponding parameters setup, the reader is referred to [13]. Correspondingly, we define the corresponding dual equation of (1.1)

$$YBY - YA - DY + C = 0. \tag{1.4}$$

To facilitate our discussion, we need a nonsingular M-matrix or a singular irreducible M-matrix given by

$$M = \begin{bmatrix} D & -C \\ -B & A \end{bmatrix} \tag{1.5}$$

and its relative matrix

$$H = JM, \tag{1.6}$$

where $J = \mathrm{diag}(I_n, -I_n)$ with I_n to be the n by n identity matrix. Our interest in this study is to find the minimal nonnegative solution X of (1.1). The existence conditions of the minimal nonnegative solution are shown by Juang et al. in [13]. Iterative methods for solving this problem are numerous and can be divided into two major categories.

One is the method sharing a computational cost of $O(n^2)$ arithmetic operations (ops) per step, but converges linear or sublinear. The representative method of the first category is the *simple iteration method* (SI) or *vector iteration method*, which is first proposed by Lu [15]. This method is very simple and requires a computational cost of $4n^2$ ops per step. Recently, three more methods, modified simple iteration (MSI), nonlinear block Jacobi method (NBJ) and the nonlinear block Gass-Seidel method (NBGS), based on Lu's method are proposed in [1, 2]. It has been shown in [10] that if $(\alpha, c) \neq (0, 1)$, the speed of convergence of the NBGS is faster than the other three. Generally speaking, the iterative methods mentioned above can be classified as accelerated variants of the well-known fixed-point iterations. Also, in [10] we know that all these four methods can provide a linear convergence, if $(\alpha, c) \neq (0, 1)$ and a sublinear convergence, if $(\alpha, c) = (0, 1)$.

The other is a method with a cost of $O(n^3)$ ops but provides quadratic convergence. Despite of the complexity, quadratically convergent methods are much to be desired in practice. There are several good algorithms that can cause quadratic convergence, for example, the Newton method [9, 4] and the structure-preserving doubling algorithm (SDA) [11, 8]. However, when $(\alpha, c) = (0, 1)$, both Newton method and the SDA algorithm are not always valid and require special attention.

In this work we fine-tune the customary SDA algorithm and make it always workable and quadratical convergent when solving (1.1). The SDA algorithm was first proposed by Guo et.al. [11] for solving the NARE. In [11, 5], it has been shown to be quadratically convergent, if $(\alpha, c) \neq (0, 1)$ and linearly convergent with rate $1/2$, if $(\alpha, c) = (0, 1)$. The later case is the so-called "critical case" and is the most challenging problem that we will encounter when solving (1.1). Roughly speaking, the critical case embedded with some type of singularity, i.e., the matrix H has two zero eigenvalues, that will significantly reduce the speed of convergence. In [8], Guo et al. propose an efficient method based on a single-shift technique to accelerate the computation of the minimal nonnegative solution so that one singularity can be removed. They also show that the doubling algorithm applied to the shifted equation of (1.1) converges faster than the doubling algorithm applied to (1.1), if no breakdown occurs. The approach of removing two zero eigenvalues of H has also been introduced in [8], but again the convergence of the doubling algorithm cannot be guaranteed. Our contribution in this paper, which we think is new in theory, is to provide a detailed analysis of changes in the eigenvalue distribution of matrices H and M as the shift procedures are employed. Through this discussion, the quadratic convergence of the SDA is guaranteed via the duble-shift technique to remove two singularities. Most important of all, the minimal nonnegative solution of the duble-shift model is shown to be equal to that of the original model. We believe such results are the first detailed proofs of the eigenvalue analysis of H and M and their corresponding matrices with shift procedures and should be of great significance for solving the NARE.

The organization of this paper is as follows. In Section 2, we review some of the main results and definitions that will be used for subsequent discussion. In Section 3, we provide a complete discussion on the shifted modifications for the SDA algorithm. We show that the SDA algorithm applied to the double-shift problem is always accessible and the solution obtained from the double-shift problem is equal to the original NARE problem. In Section 4, advantages of the shifting

technique applied to the SI algorithm have been thoroughly investigated. In Section 5, we present a few numerical experiments to show the practicability and effectiveness of the shifting procedure and concluding remarks are given in Section 6.

2. Preliminaries. In this section we briefly review the definitions of Z-matrix and M-matrix and discuss further some of their properties which are required in the statements and in the proofs discussed in the following sections. We also summarize the popular algorithm, SDA, for our numerical experiments as we shall see below.

2.1. Definition and Theorems. In order to formalize our discussion, we start by introducing the following two definitions.

DEFINITION 2.1. *A matrix $A = (a_{ij}) \in \mathbb{R}^{n \times n}$ is called a* Z-matrix *if $a_{ij} \leq 0$ for all $i \neq j$.* Note that for any Z-matrix A, there exists a matrix $B \in \mathbb{R}^{n \times n}$ with $B \geq 0$ and some $\alpha \in \mathbb{R}$ such that $A = \alpha I - B$ where I is the identity matrix. Also, the definition of Z-matrix plays an important role in defining a given matrix to be an *M-matrix*.

DEFINITION 2.2. *A Z-matrix A is called an* M-matrix *if $A = \alpha I - B$ with $B \geq 0$ and $\alpha \geq \rho(B)$, where $\rho(B)$ is the spectral radius of B. It is called a singular M-matrix if $\alpha = \rho(B)$ and a nonsingular M-matrix if $\alpha > \rho(B)$.*

There are a great many different conditions, which are mathematically intriguing and important for applications, that discuss the necessary and sufficient conditions for a given Z-matrix to be an M-matrix. For our subsequent discussions, we apply the following two well known and useful results in the study of M-matrices.

THEOREM 2.3. *[3] If $A \in \mathbb{R}^{n \times n}$ is a Z-matrix, the following statements are equivalent:*

1. A is a nonsingular M-matrix.

2. $\sigma(A) \subset \mathbb{C}_+$.

3. $Av > 0$ holds for some positive vector $v \in \mathbb{R}^n$.

4. $A^{-1} \geq 0$.

THEOREM 2.4. *[9] If the matrix (1.5) is a nonsingular M-matrix, then the NARE (1.1) and its dual equation (1.4) have minimal nonnegative solutions X and Y, respectively. Also, matrices $D - CX$ and $A - BY$ are nonsingular M-matrix.*

Note that the conditions we list here are only a selection from many more useful ones. See [3, 9, 12, 16] for a longer list of conditions and references to the proofs.

2.2. SDA Algorithm. In [11], Guo et al. come up with the SDA algorithm for solving NARE problems and show that if the matrix M (1.5) is a nonsingular M-matrix (irreducible singular M-matrix [5]), the SDA algorithm is well-defined and quadratically convergent (at least linearly convergent with rate $1/2$). Its idea is based on the doubling transformation. For more details of the doubling transformation, the reader is referred to [11, Theorem 2.1]. The SDA algorithm starts by choosing a positive scalar γ with

$$\gamma \geq \max \left\{ \max_{1 \leq i \leq n} a_{ii}, \max_{1 \leq i \leq n} d_{ii} \right\}.$$

Let

$$E_0 = I_n - 2\gamma V_\gamma^{-1}, \quad F_0 = I_n - 2\gamma W_\gamma^{-1},$$
$$G_0 = 2\gamma D_\gamma^{-1} C W_\gamma^{-1}, \quad H_0 = 2\gamma W_\gamma^{-1} B D_\gamma^{-1},$$

where

$$A_\gamma = A + \gamma I_n, \quad D_\gamma = D + \gamma I_n,$$
$$W_\gamma = A_\gamma - B D_\gamma^{-1} C, \quad V_\gamma = D_\gamma - C A_\gamma^{-1} B.$$

Then, the SDA algorithm presented in [11] is given by

$$E_{k+1} = E_k(I_n - G_kH_k)^{-1}E_k, \tag{2.3a}$$

$$F_{k+1} = F_k(I_n - H_kG_k)^{-1}F_k, \tag{2.3b}$$

$$G_{k+1} = G_k + E_k(I_n - G_kH_k)^{-1}G_kF_k, \tag{2.3c}$$

$$H_{k+1} = H_k + F_k(I_n - H_kG_k)^{-1}H_kE_k, \tag{2.3d}$$

where the sequence H_k and G_k will converge to the minimal nonnegative solutions X of (1.1) and Y of (1.4) quadratically.

2.3. Spectrum Analysis. Recall that in the critical case $(\alpha, c) = (0, 1)$, the matrix M (1.5) is an irreducible singular M-matrix [6] and the corresponding matrix H (1.6) has a double zero eigenvalue with the geometric multiplicity equal to one. To be specific, the matrix H has $2n$ real eigenvalues $\nu_n, \ldots, \nu_1, \lambda_1, \ldots, \lambda_n$, which satisfy the following order [13]:

$$\frac{-1}{\omega_n} < \nu_n < \frac{-1}{\omega_{n-1}} < \ldots < \frac{-1}{\omega_2} < \nu_2 < \frac{-1}{\omega_1} < \nu_1 = 0 = \lambda_1 < \frac{1}{\omega_1} < \lambda_2 < \frac{1}{\omega_2} < \ldots < \lambda_n < \frac{1}{\omega_n}. \tag{2.4}$$

The phenomenon is called *eigenvalue interlacing*. Moreover,

$$\sigma(D - CX) = \{\lambda_2, \ldots, \lambda_n, 0\}, \tag{2.5a}$$

$$\sigma(A - BY) = \{0, -\mu_1, \ldots, -\mu_n\}, \tag{2.5b}$$

if X and Y are the minimal nonnegative solutions of (1.1) and (1.4), respectively [6]. Paralleling the above distribution, the following theorem shows that all eigenvalues of M are real and nonnegative. In fact, M has n specific eigenvalues $\frac{1}{\omega_i}$, for $i = 1, \ldots, n$.

THEOREM 2.5. *Let M be the matrix defined in (1.5) with $(\alpha, c) = (0, 1)$. Then M has $2n$ real eigenvalues, where one part of the eigenvalues of M are $0, \frac{1}{\omega_1}, \ldots, \frac{1}{\omega_n}$ and the others are $\mu_1, \ldots, \mu_{n-1}$ such that the eigenvalues can be arranged in the following order:*

$$0 < \frac{1}{\omega_1} < \mu_1 < \frac{1}{\omega_2} < \mu_2 < \cdots < \mu_{n-1} < \frac{1}{\omega_n}.$$

Proof. Consider the characteristic polynomial of M defined by

$$f(\lambda) \equiv \det(M - \lambda I_n) = \det\left(\begin{bmatrix} \Gamma - \lambda I_n & \\ & \Delta - \lambda I_n \end{bmatrix} - \begin{bmatrix} q \\ e \end{bmatrix} \begin{bmatrix} e^\top & q^\top \end{bmatrix}\right)$$

$$= \det\left(\begin{bmatrix} (\Gamma - \lambda I_n) & \\ & (\Delta - \lambda I_n) \end{bmatrix}\right)(1 - \begin{bmatrix} e^\top & q^\top \end{bmatrix}\begin{bmatrix} (\Gamma - \lambda I_n)^{-1} & \\ & (\Delta - \lambda I_n)^{-1} \end{bmatrix}\begin{bmatrix} q \\ e \end{bmatrix})$$

$$= \prod_{1 \le i \le n}(\gamma_i - \lambda)(\delta_i - \lambda)(1 - \sum_{1 \le j \le n}(\frac{q_j}{\gamma_j - \lambda} + \frac{q_j}{\delta_j - \lambda})) \tag{2.6}$$

The last equation (2.6) is called the secular equation of $M - \lambda I$. Notice that $\gamma_i = \delta_i = \frac{1}{\omega_i}$, $q_i = \frac{c_i}{2\omega_i}$ for $1 \le i \le n$ when $(\alpha, c) = (0, 1)$. Thus, through a straightforward calculation, we have

$$f(\lambda) = \prod_{1 \le i \le n}(\frac{1}{\omega_i} - \lambda)^2(1 - \sum_{1 \le j \le n}\frac{c_j}{(1 - \omega_j\lambda)})$$

$$= -\prod_{1 \le i \le n}(\frac{1}{\omega_i} - \lambda)^2 \sum_{1 \le j \le n}\frac{c_j\lambda}{\frac{1}{\omega_j} - \lambda}$$

$$= -\lambda \prod_{1 \le i \le n}(\frac{1}{\omega_i} - \lambda)\left(\sum_{1 \le j \le n} c_j \prod_{k \ne j, 1 \le k \le n}(\frac{1}{\omega_k} - \lambda)\right).$$

Thus, f has roots $0, \frac{1}{\omega_1}, \cdots, \frac{1}{\omega_n}$. To complete the proof of the theorem, let

$$g(\lambda) = \sum_{1 \le j \le n} c_j \prod_{k \ne j, 1 \le k \le n} (\frac{1}{\omega_k} - \lambda).$$

The sign of $g(\frac{1}{\omega_j})$ is $(-1)^{j-1}$ since the monotonicity of $\{\omega_j\}$, the intermediate value theorem indicates that g has at least roots in $(\frac{1}{\omega_j}, \frac{1}{\omega_{j+1}})$ for $1 \le j \le n-1$. Together with the fact that the degree of g is $n - 1$. The proof of the theorem is thus complete. $\square$

It should be noted that

$$H \begin{bmatrix} I_n \\ X \end{bmatrix} = \begin{bmatrix} I_n \\ X \end{bmatrix} (D - CX). \tag{2.7}$$

From the above theorem and (2.7), we know that the minimal nonnegative solution X is related to an invariant subspace with nonnegative eigenvalues of H. Also, it is clear that $q^\top \Gamma^{-1} e + e^\top \Delta^{-1} q = c = 1$. We then have the fact [4] that the matrix H has *a right eigenvector* $v^\top = [v_1^\top, v_2^\top]$, with $v_1 = \Gamma^{-1} q$ and $v_2 = \Delta^{-1} e$, so that

$$Hv = 0. \tag{2.8}$$

By applying this right eigenvector v, *a left eigenvector* $u^\top = [u_1^\top, u_2^\top]$, with $u_1 = \Gamma^{-1} e$ and $u_2 = -\Delta^{-1} q$ of H, corresponding to the eigenvalue 0, can be obtained without any trouble by directly checking that

$$u^\top H = 0. \tag{2.9}$$

Corresponding to the matrix H, the matrix $M = JH$ has the right and left eigenvectors v and $u^\top J$. Also, it can be seen that $u_1^\top v_1 + u_2^\top v_2 = 0$. Applying the eigenpair information, we have the following important result given in [6, 8].

THEOREM 2.6. *Let M be an irreducible singular M-matrix as defined in (1.5), and let X and Y are the minimal nonnegative solutions of (1.1) and (1.4), respectively. Suppose that corresponding to the zero eigenvalue, the right and left eigenvectors of M are $v^\top = [v_1^\top, v_2^\top]$ and $u^\top = [u_1^\top, -u_2^\top]$. If $(\alpha, c) = (0, 1)$, then the following properties are satisfied:*

$$Xv_1 = v_2, \quad u_2^\top X = -u_1^\top, \quad and \quad Yv_2 = v_1. \tag{2.10}$$

It was shown in [8], that the matrix X is the minimal nonnegative solution of 1.1 if and only if $X^\top$ is the minimal nonnegative solution of the equation

$$X^\top C^\top X^\top - X^\top A^\top - D^\top X^\top + B^\top = 0. \tag{2.11}$$

The same statement can be applied to the dual equation (1.4). Its proof is simply based on taking the transpose on both sides of (1.1).

COROLLARY 2.7. *The matrix Y is the minimal nonnegative solution of (1.4) if and only if $Y^\top$ is the minimal nonnegative solution of the equation*

$$Y^\top B^\top Y^\top - Y^\top D^\top - A^\top Y^\top + C^\top = 0. \tag{2.12}$$

Following Corollary (2.7), we want to know that whether there exists a relationship between the left eigenvector of M and the minimal nonnegative solution Y. To begin with, let

$$M_t = \begin{bmatrix} D^\top & -B^\top \\ -C^\top & A^\top \end{bmatrix}, \tag{2.13}$$

be the corresponding M-matrix of (2.12). Note that M_t has a right eigenvector $[u_1^\top, -u_2^\top]^\top$ and a left eigenvector $[v_1^\top, v_2^\top]$ corresponding to the eigenvalue 0. Equipped with the notations given in (1.2), the matrix M_t is again an irreducible singular M-matrix if $(\alpha, c) = (0, 1)$. Then, Theorem 2.6 asserts that $Y^\top u_1 = -u_2$. Namely, we have derived the following important relationship between the left eigenvector u and the minimal solution Y,

$$u_1^\top Y = -u_2^\top. \tag{2.14}$$

On the other hand, we know that the convergence rate of the SDA algorithm is determined by

$$\limsup_{k \to \infty} {}^{2^k}\!\sqrt{\|H_k - X\|} \leq \rho(C_\gamma(D - CX))\rho(C_\gamma(A - BY)), \tag{2.15a}$$

$$\limsup_{k \to \infty} {}^{2^k}\!\sqrt{\|G_k - Y\|} \leq \rho(C_\gamma(D - CX))\rho(C_\gamma(A - BY)), \tag{2.15b}$$

where

$$C_\gamma : z \to \frac{z - \gamma}{z + \gamma} \tag{2.16}$$

is the Cayley transform and the scalar $\gamma > 0$ [8]. Note that from (2.5), we have $\rho(C_\gamma(D - CX) = \rho(C_\gamma(A - BY)) = 1$. It follows that no further conclusion of the convergence rate of the SDA algorithm can be derived from the fact (2.15) except that the linear convergence is guaranteed. In the subsequent section, we want to know that how the shift procedures affect the convergence rate.

3. Properties of the Shifted NARE. In this section, a detailed analysis of the eigenvalue distribution of the matrix M is provided with respect to the the critical case $(\alpha, c) = (0, 1)$. It is shown that under the shifting technique, the matrix M is still an M-matrix and the SDA algorithm is guaranteed to converge. The minimal nonnegative solution in the shifted NARE problems are proved to be equal to the minimal nonnegative solution of (1.1). Last but not least, the SDA algorithm is shown to be accelerated by removing the singularities embedded in the matrix H.

3.1. Single Shift. Let $\widehat{H}$ be the rank-one modification of the matrix H which is defined by

$$\widehat{H} = H + \eta v r^\top, \tag{3.1}$$

where $\eta > 0$ is a scalar and $r \geq 0$ is a vector satsifying $r^\top v = 1$. To be specific, we write $r^\top = [r_1^\top, r_2^\top]$, where $r_1 = e$, $r_2 = q$. Then, two matrices $\widehat{H}$ and $\widehat{M}$ are denoted by

$$\widehat{H} = \begin{bmatrix} \widehat{D} & -\widehat{C} \\ \widehat{B} & -\widehat{A} \end{bmatrix}, \quad \widehat{M} = \begin{bmatrix} \widehat{D} & -\widehat{C} \\ -\widehat{B} & \widehat{A} \end{bmatrix}, \tag{3.2}$$

where

$$\widehat{D} = D + \eta v_1 r_1^\top, \quad \widehat{C} = C - \eta v_1 r_2^\top,$$
$$\widehat{B} = B + \eta v_2 r_1^\top, \quad \widehat{A} = A - \eta v_2 r_2^\top. \tag{3.3}$$

It follows from the specific structure of $\widehat{M}$ given in (3.2) that the matrix $\widehat{M}$ is irreducible. The nice feature of this rank-one modification is that one zero eigenvalue of H will be replaced by the scalar $\eta > 0$. This can be seen by directly applying the following useful lemma shown in [8].

LEMMA 3.1. *Let T be a singular matrix with $Tv = 0$ for some nonzero vector v. If r is a vector so that $r^\top v = 1$, then for any scalar r, the eigenvalues of the matrix*

$$\widehat{T} = T + \eta v r^\top,$$

consist of those of T, except that one zero eigenvalue of T is replaced by η.

It can be seen that from Lemma 3.1 the eigenvalues of H and $\widehat{H}$ are the same except that one zero eigenvalue is shifted to η. In the next theorem, we want to show that despite of the rank one modification, the eigenvalues of $\widehat{M}$ are equal to those of M.

COROLLARY 3.2. *Let M and $\widehat{M}$ be defined in (1.5) and (3.2), respectively. Then, the characteristic polynomials of M and $\widehat{M}$ are conincident. That is, the eigenvalues of M and $\widehat{M}$ are equal.*

Proof. This proof can be easily obtained by studying the characteristic polynomial of $\widehat{M}$. We know that the characteristic polynomial of $\widehat{M}$, denoted by $\widehat{f}(\lambda)$, is defined by

$$\widehat{f}(\lambda) \equiv \det(\widehat{M} - \lambda I_{2n}) = \det(\begin{bmatrix} \Gamma - \lambda I_n & \\ & \Delta - \lambda I_n \end{bmatrix} + \begin{bmatrix} (-I_n + \eta\Gamma^{-1})q \\ (-I_n - \eta\Delta^{-1})e \end{bmatrix} \begin{bmatrix} e^\top & q^\top \end{bmatrix})$$

$$= \prod_{i=1}^{n}(\frac{1}{\omega_i} - \lambda)^2 \det(1 + \begin{bmatrix} e^\top & q^\top \end{bmatrix} \begin{bmatrix} (\Gamma - \lambda I_n)^{-1} & \\ & (\Delta - \lambda I_n)^{-1} \end{bmatrix} \begin{bmatrix} (-I_n + \eta\Gamma^{-1})q \\ (-I_n - \eta\Delta^{-1})e \end{bmatrix})$$

$$= -\prod_{i=1}^{n}(\frac{1}{\omega_i} - \lambda)^2 \sum_{j=1}^{n} \frac{c_j \lambda}{\frac{1}{\omega_j} - \lambda}. \tag{3.4}$$

From (3.4), we know that the eigenvalues of $\widehat{M}$ are precisely those of M. $\blacksquare$

THEOREM 3.3. *The matrix $\widehat{M}$ defined by equation (3.2) is a Z-matrix if and only if the parameter η, defined in (3.9) satisfy*

$$0 < \eta \le \frac{1}{\omega_1}. \tag{3.5}$$

Proof. From (3.2), $\widehat{M}$ is a Z-matrix if and only if $\widehat{B} \ge 0$, $\widehat{C} \ge 0$, and $\widehat{D}$ and $\widehat{A}$ are Z-matrices. Note that

$$\begin{aligned} \widehat{D} &= \Gamma + (-I_n + \eta\Gamma^{-1})qe^\top, & \widehat{C} &= (I_n - \eta\Gamma^{-1})qq^\top, \\ \widehat{B} &= (I + \eta\Delta^{-1})ee^\top > 0, & \widehat{A} &= \Delta + (-I_n - \eta\Delta^{-1})eq^\top. \end{aligned} \tag{3.6}$$

The sufficient and necessary condition such that the matrix $\widehat{M}$ is a Z-matrix is that $\widehat{C} \ge 0$, and $\widehat{D}$ and $\widehat{A}$ are Z-matrices. This implies that

$$-1 + \eta\omega_1 \le 0. \tag{3.7}$$

Since η is positive, we have the fact that $\widehat{M}$ is a Z-matrix if and only if (3.5) is satisfied. $\blacksquare$

Using Corollary 3.2 and the given constraint (3.5) in Theorem 3.3, we know that $\widehat{M}$ is an irreducible M-matrix and the SDA algorithm is guaranteed to be applicable. It is known that the minimal nonnegative solution $\widehat{X}$ of the single shifted NARE is equivalent to the minimal nonnegative solution X of (1.1) [8]. Thus, we have

$$\limsup_{k\to\infty} \sqrt[2^k]{\|H_k - X\|} \le \rho(C_\gamma(\widehat{D} - \widehat{C}X))\rho(C_\gamma(\widehat{A} - \widehat{B}\widehat{Y})) < 1, \tag{3.8}$$

since $\rho(C_\gamma(\widehat{D} - \widehat{C}X) < 1$ and $\rho(C_\gamma(\widehat{A} - \widehat{B}\widehat{Y})) = 1$. It concludes that the convergence of the SDA algorithm with a single shift is faster than that with no shift. Based on all the properties stated above, it is illuminating to begin the analysis of the double shifting technique.

3.2. Double Shift. In order to remove all zero eigenvalues of H, we define the double shifted matrix $\overline{H}$,

$$\overline{H} = H + \eta v r^\top + \xi s u^\top = \begin{bmatrix} \overline{D} & -\overline{C} \\ \overline{B} & -\overline{A} \end{bmatrix}, \tag{3.9}$$

where $\eta > 0$, $\xi < 0$, $p^\top$ and $q^\top$ such that $p^\top v = q^\top u = 1$, each size of sub-matrices $\overline{A}, \overline{B}, \overline{C}$ and $\overline{D}$ are n square. This is the so called *double shifting* technique. Indeed, it can be seen that if we choose $s^\top = [s_1^\top, s_2^\top]$ with $s_1 = q$ and $s_2 = -e$ and the same vectors r, u and v as defined above, then the vectors p and q satisfy the fact that

$$r^\top v = s^\top u = e^\top \Gamma^{-1} q + q^\top \Delta^{-1} e = 1. \tag{3.10}$$

From Lemma 3.1, we know that the double shifting technique will move one zero eigenvalue of H to $\eta > 0$ and the other to $\xi < 0$ and keep the nonzero eigenvalues unchanged. With this in mind, the shift technique introduced in formula (3.9) will make the new matrix $\overline{H}$ nonsingular. Also, we can define a duble shifted NARE in $\overline{X} \in \mathbb{R}^{n \times n}$ associate with the matrix $\overline{H}$ as follows:

$$\overline{X}\,\overline{C}\,\overline{X} - \overline{X}\,\overline{D} - \overline{A}\,\overline{X} + \overline{B} = 0, \tag{3.11a}$$

and the dual duble shifted NARE in $\overline{Y} \in \mathbb{R}^{n \times n}$

$$\overline{Y}\,\overline{B}\,\overline{Y} - \overline{Y}\,\overline{A} - \overline{D}\,\overline{Y} + \overline{C} = 0, \tag{3.11b}$$

where

$$\begin{aligned}
\overline{D} &= D + \eta v_1 r_1^\top + \xi s_1 u_1^\top, & \overline{C} &= C - \eta v_1 r_2^\top - \xi s_1 u_2^\top, \\
\overline{B} &= B + \eta v_2 r_1^\top + \xi s_2 u_1^\top, & \overline{A} &= A - \eta v_2 r_2^\top - \xi s_2 u_2^\top.
\end{aligned} \tag{3.12}$$

In what follows, we show that under suitable assumptions on parameters η and ξ, the matrix $\overline{M}$ defined by

$$\overline{M} = \begin{bmatrix} \overline{D} & -\overline{C} \\ -\overline{B} & \overline{A} \end{bmatrix}, \tag{3.13}$$

is a nonsingular M-matrix, that is, the SDA algorithm is well-defined and applicable to the NARE (3.11a). We start our proof by showing that this matrix $\overline{M}$ is a Z-matrix for some parameters η and ξ.

THEOREM 3.4. *The matrix $\overline{M}$ defined by equation (3.13) is a Z-matrix if and only if the parameters, η and ξ, defined in (3.9) satisfy the following two conditions:*

$$0 < \eta < \frac{1}{\omega_1}, \tag{3.14a}$$

$$\frac{-1 + \eta \omega_1}{\omega_1} \le \xi < 0. \tag{3.14b}$$

Proof. It follows from (3.13) we know that $\overline{M}$ is a Z-matrix if and only if $\overline{B} \ge 0$, $\overline{C} \ge 0$, and $\overline{D}$ and $\overline{A}$ are Z-matrices. Also, from (3.12) we have

$$\begin{aligned}
\overline{D} &= \Gamma + (-I_n + \eta \Gamma^{-1}) q e^\top + \xi q e^\top \Gamma^{-1}, \\
\overline{C} &= (I_n - \eta \Gamma^{-1}) q q^\top + \xi q q^\top \Delta^{-1}, \\
\overline{B} &= (I + \eta \Delta^{-1}) e e^\top - \xi e e^\top \Gamma^{-1} > 0, \\
\overline{A} &= \Delta + (-I_n - \eta \Delta^{-1}) e q^\top - \xi e q^\top \Delta^{-1}.
\end{aligned}$$

Therefore, in order to get a Z-matrix $\overline{M}$, we only need to consider when $\overline{C} \geq 0$, and $\overline{D}$ and $\overline{A}$ are Z-matrices. This gives rise to the following three sufficient and necessary conditions:

$$\begin{cases} -1 + \eta\omega_1 + \xi\omega_n & \leq & 0, \\ -1 + \eta\omega_1 - \xi\omega_1 & \leq & 0, \\ -1 - \eta\omega_n - \xi\omega_1 & \leq & 0. \end{cases} \tag{3.15}$$

It follows from (3.15) and the initial conditions $\eta > 0$ and $\xi < 0$ that $\overline{M}$ is a Z-matrix if and only if (3.14a) and (3.14b) are satisfied. $\square$

To simplify our discussion, we define

$$\Omega = \{(\eta, \xi); 0 < \eta < \frac{1}{\omega_1}, \frac{-1 + \eta\omega_1}{\omega_1} \leq \xi < 0\}. \tag{3.16}$$

Our next approach is to show that the matrix $\overline{M}$ is indeed an M-matrix. That is, the iterative processes in SDA algorithm do not break down and convergence quadratically. To begin with, we introduce the following two lemmas.

LEMMA 3.5. Let c_i and ω_i, for $i = 1, \ldots, n$, be defined in (1.1). Given $\lambda \in \mathbb{R}$ and $\lambda \neq \frac{1}{\omega_i}$, for $i = 1, \ldots, n$, we define

$$g_1(\lambda) = \lambda \sum_{i=1}^{n} \frac{c_i}{\frac{1}{\omega_i} - \lambda}, \quad g_2(\lambda) = \sum_{i=1}^{n} \frac{c_i\omega_i}{\frac{1}{\omega_i} - \lambda}, \quad g_3(\lambda) = \sum_{i=1}^{n} \frac{c_i}{\omega_i(\frac{1}{\omega_i} - \lambda)}. \tag{3.17}$$

Then, the following properties hold:

1. $g_1(\lambda) - \lambda^2 g_2(\lambda) = \lambda \sum_{i=1}^{n} c_i\omega_i$.
2. $g_1(\lambda) - g_3(\lambda) = -1$.
3. If $\lambda \in (\frac{1}{\omega_k}, \frac{1}{\omega_{k+1}})$, then $g_3(\lambda) \geq g_1(\lambda)\frac{1}{\lambda\omega_k} \geq g_2(\lambda)\frac{1}{\omega_k}$.

Proof. The first two properties are following from the direct computation. To see this, applying the conditions in (3.17), we have

$$g_1(\lambda) - \lambda^2 g_2(\lambda) = \lambda \sum_{i=1}^{n} \frac{(c_i - \lambda c_i\omega_i)\omega_i}{\omega_i(\frac{1}{\omega_i} - \lambda)}$$

$$= \lambda \sum_{i=1}^{n} c_i\omega_i.$$

$$g_1(\lambda) - g_3(\lambda) = \sum_{i=1}^{n} \frac{(\lambda c_i\omega_i - c_i)}{\omega_i(\frac{1}{\omega_i} - \lambda)} = -1.$$

Using the triangle inequality and $\lambda \in (\frac{1}{\omega_k}, \frac{1}{\omega_{k+1}})$, we obtain

$$
\begin{aligned}
g_3(\lambda) &= \sum_{i=1}^{n} \frac{c_i}{\omega_i(\frac{1}{\omega_i} - \lambda)} \\
&\geq \sum_{i=1}^{k} \frac{c_i}{\omega_k(\frac{1}{\omega_i} - \lambda)} + \sum_{i=k+1}^{n} \frac{c_i}{\omega_{k+1}(\frac{1}{\omega_i} - \lambda)} \\
&\geq g_1(\lambda)\frac{1}{\lambda\omega_k} \\
&\geq \left(\sum_{i=1}^{k} \frac{c_i\omega_k}{(\frac{1}{\omega_i} - \lambda)} + \sum_{i=k+1}^{n} \frac{c_i\omega_{k+1}}{(\frac{1}{\omega_i} - \lambda)} \right) \frac{1}{\omega_k} \\
&\geq g_2(\lambda)\frac{1}{\omega_k}.
\end{aligned}
$$

$\square$

We have now seen that the relationships among $g_1(\lambda)$, $g_2(\lambda)$, and $g_3(\lambda)$. Let $g(\lambda)$ to be a function satisfying

$$g(\lambda) \equiv \lambda g_1(\lambda) + \eta\xi g_2(\lambda)g_3(\lambda), \tag{3.18}$$

where $(\eta, \xi) \in \Omega$. Our next approach is to show that for each subinterval $(\frac{1}{\omega_k}, \frac{1}{\omega_{k+1}})$ with $k = 1, \ldots, n-1$, there exists a point λ so that $g(\lambda) > 0$. This property is a stepping stone for showing that $\overline{M}$ is an M-matrix.

LEMMA 3.6. *Let c_i and ω_i, for $i = 1, \ldots, n$, be defined in (1.1). It then follows that there exists a point $\lambda_k \in (\frac{1}{\omega_k}, \frac{1}{\omega_{k+1}})$, for $k = 1, \ldots, n-1$, so that the function $g(\lambda)$ of (3.18) is greater than zero.*

Proof. Note that $g_3(\lambda)$ is a continuous function on $(\frac{1}{\omega_k}, \frac{1}{\omega_{k+1}})$, $\lim_{\lambda \to \frac{1}{\omega_k}^+} g_3(\lambda) = -\infty$, and $\lim_{\lambda \to \frac{1}{\omega_{k+1}}^-} g_3(\lambda) = +\infty$, for all $k = 1, \ldots, n-1$. Thus, there exists a point $\lambda_k \in (\frac{1}{\omega_k}, \frac{1}{\omega_{k+1}})$ such that

$$g_3(\lambda_k) = \frac{4\omega_1^2}{\omega_k\omega_{k+1}}. \tag{3.19}$$

Since $\omega_1 > \omega_2 > \cdots > \omega_n$, we have the fact that $g_3(\lambda_k) > 4$. It follows from Lemma 3.5 that $g_1(\lambda_k) > 0$.

We first assume that $g_2(\lambda_k) < 0$ for this specific λ_k, then it is clear that $g(\lambda_k) = \lambda_k g_1(\lambda_k) + \eta\xi g_2(\lambda_k)g_3(\lambda_k) > 0$, since $\eta\xi < 0$. We now assume that $g_2(\lambda_k) > 0$. Combining the inequalities (3.14a) with (3.14b), we have

$$-\frac{1}{4\omega_1^2} \leq \eta\xi < 0. \tag{3.20}$$

Then, by (3.18) we get

$$
\begin{aligned}
g(\lambda_k) &\geq \sum_{i=1}^{n} \frac{c_i(\lambda_k - \frac{\omega_i}{\omega_k\omega_{k+1}})}{\frac{1}{\omega_i} - \lambda_k} \\
&= \sum_{i=1}^{k} \frac{c_i(\lambda_k - \frac{\omega_i}{\omega_k\omega_{k+1}})}{\frac{1}{\omega_i} - \lambda_k} + \sum_{i=k+1}^{n} \frac{c_i(\lambda_k - \frac{\omega_i}{\omega_k\omega_{k+1}})}{\frac{1}{\omega_i} - \lambda_k} \geq 0, \tag{3.21}
\end{aligned}
$$

since $\lambda_k - \frac{\omega_i}{\omega_k\omega_{k+1}} < \frac{1}{\omega_{k+1}} - \frac{\omega_k}{\omega_k\omega_{k+1}} = 0$, for $1 \le i \le k$, and $\lambda_k - \frac{\omega_i}{\omega_k\omega_{k+1}} > \omega_k - \frac{\omega_{k+1}}{\omega_k\omega_{k+1}} = 0$, for $k+1 \le i \le n$.
□

Now we have enough tools to validate that the given matrix $\overline{M}$ is indeed an M-matrix. In particular, we can also dig out the eigenvalue distribution of matrix $\overline{M}$.

THEOREM 3.7. *If $(\eta, \xi) \in \Omega$, then the matrix $\overline{M}$ defined by equation (3.13) is an M-matrix. In particular, $\overline{M}$ has $2n$ positive real eigenvalues $\overline{\lambda}_1, \ldots, \overline{\lambda}_{2n}$ satisfying*

$$0 < \overline{\lambda}_1 < \overline{\lambda}_2 < \frac{1}{\omega_1} < \overline{\lambda}_3 < \overline{\lambda}_4 < \frac{1}{\omega_2} < \cdots < \frac{1}{\omega_{n-1}} < \overline{\lambda}_{2n-1} < \overline{\lambda}_{2n} < \frac{1}{\omega_n} \qquad (3.22)$$

Proof. Since the matrix $\overline{H}$ of (3.9) is nonsingular, it is clear that $\overline{M} = J\overline{H}$ is nonsingular. Also, Theorem 3.4 implies $\overline{M}$ is a Z-matrix. In order to show that $\overline{M}$ is an M-matrix, it suffices to show that all eigenvalues of $\overline{M}$ have positive real part. Indeed, all eigenvalues of $\overline{M}$ are positive real numbers and satisfy the interlacing property.

We first consider the characteristic polynomial $\bar{f}(\lambda)$ of $\overline{M}$ defined by

$$
\begin{aligned}
\bar{f}(\lambda) &\equiv \det(\overline{M} - \lambda I_{2n}) \\
&= \det\left(\begin{bmatrix} \Gamma - \lambda I_n & \\ & \Delta - \lambda I_n \end{bmatrix} + \begin{bmatrix} (-I_n + \eta\Gamma^{-1})q & q \\ (-I_n - \eta\Delta^{-1})e & e \end{bmatrix} \begin{bmatrix} e^\top & q^\top \\ \xi e^\top \Gamma^{-1} & -\xi q^\top \Delta^{-1} \end{bmatrix}\right) \\
&= -\prod_{i=1}^{n}\left(\frac{1}{\omega_i} - \lambda\right)^2 g(\lambda) \qquad (3.23) \\
&= -\sum_{i=1}^{n} c_j \prod_{1 \le s \le n, s \ne i}\left(\frac{1}{\omega_i} - \lambda\right) \sum_{k=1}^{n} c_k \left[\left(-\lambda^2 + \frac{\lambda}{\omega_k} + \frac{\xi\eta\omega_k}{\omega_j}\right) \prod_{1 \le s \le n, s \ne k}\left(\frac{1}{\omega_s} - \lambda\right)\right], \quad (3.24)
\end{aligned}
$$

where $g(\lambda)$ is the function given in (3.18). By direct substitution of $\frac{1}{\omega_k}$ in (3.24), we have $\bar{f}(\frac{1}{\omega_k}) > 0$, for $k = 1, \ldots, n$. Also, it follows from (3.23) that $\bar{f}(0) > 0$. If we can find a point λ satisfying $\bar{f}(\omega) < 0$ in each subinterval $(\frac{1}{\omega_k}, \frac{1}{\omega_{k+1}})$, for $k = 1, \ldots, n$ and the interval $(0, \frac{1}{\omega_1})$, then the intermediate value theorem imply that the distribution of eigenvalues of $\overline{M}$ arranged in (3.22) is valid. This also gives rise to the fact that $\overline{M}$ is a nonsingular M-matrix.

Next, we consider the subinterval $(0, \frac{1}{\omega_1})$. Choosing $\lambda = \frac{1}{2\omega_1}$, it follows that

$$
\begin{aligned}
g\left(\frac{1}{2\omega_1}\right) &= \frac{1}{2\omega_1} \sum_{i=1}^{n} \frac{c_i}{\frac{1}{\omega_i} - \frac{1}{2\omega_1}} + \eta\xi \sum_{i=1}^{n} \frac{c_i\omega_i}{\frac{1}{\omega_i} - \frac{1}{2\omega_1}} \sum_{i=1}^{n} \frac{c_i}{\omega_i(\frac{1}{\omega_i} - \frac{1}{2\omega_1})} \\
&\ge \left(c_1 + \sum_{i=2}^{n} \frac{c_i\omega_i}{2\omega_1 - \omega_i}\right) - \left(c_1 + \sum_{i=2}^{n} \frac{c_i\omega_i(\frac{\omega_i}{\omega_1})}{2\omega_i - \omega_i}\right)\left(c_1 + \sum_{i=2}^{n} \frac{c_i\omega_1}{2\omega_i - \omega_i}\right) \qquad (3.25) \\
&\ge 0 \qquad (3.26)
\end{aligned}
$$

The second inequality (3.25) comes from the fact that $\eta\xi \ge -\frac{1}{4\omega_1^2}$. Also, since $\frac{\omega_k}{\omega_1} < 1$ and $\frac{c_i\omega_1}{2\omega_1 - \omega_i} < c_i$, for $i = 2, \ldots, n$, and $\sum_{i=1}^{n} c_i = 1$, we have the last inequality (3.26). For the proof of each subinterval $(\frac{1}{\omega_k}, \frac{1}{\omega_{k+1}})$, we simply apply the conclusion of Lemma 3.6. Then, (3.23) immediately implies that there exists a point $\lambda \in (\frac{1}{\omega_k}, \frac{1}{\omega_{k+1}})$ such that $f(\lambda) < 0$, for $k = 1, \ldots, n$.
□

Note that in [8] the minimal nonnegative solution X of (1.1) has been shown to be a solution of (3.11a). So far, to the best of our knowledge, no study has investigated the relation between

the solutions $\overline{X}$ and X. If there does not exist any relation between $\overline{X}$ and X, the solution obtained from the duble-shift algorithm would be exclusively meaningless. Our next result is to find this substantial link through the known fact that $\overline{M}$ is indeed an M-matrix (3.7).

THEOREM 3.8. *Let $\overline{X}$ and X be the minimal nonnegative solutions of* (3.11a) *and* (1.1), *respectively. Then,* $\sigma(\overline{D} - \overline{C}X) = \{\eta, \lambda_2, \ldots, \lambda_n\}$ *and* $\overline{X} = X$.

Proof. Let $\mathcal{R}(Z) = ZCZ - ZD - AZ + B$ and $\overline{\mathcal{R}}(Z) = Z\overline{C}Z - Z\overline{D} - \overline{A}Z + \overline{B}$. Observe first that

$$\overline{\mathcal{R}}(X) = \mathcal{R}(X) - \eta(Xv_1 - v_2)(r_2^\top X + r_1^\top) + \xi(Xs_1 - s_2)(-u_2^\top X - u_1^\top) = \mathcal{R}(X), \quad (3.27)$$

where the second equality follows directly from Theorem 2.6. This equality amounts to say that the minimal nonnegative solution of (1.1) is also a nonnegative solution of (3.11a) and the following equality is satisfied.

$$\overline{H}\begin{bmatrix} I_n \\ X \end{bmatrix} = \begin{bmatrix} I_n \\ X \end{bmatrix}(\overline{D} - \overline{C}X). \tag{3.28}$$

Recall that $u_1^\top + u_2^\top X = 0$. Then, we have

$$(\overline{D} - \overline{C}X) = D - CX + \eta v_1(r_1^\top + r_2^\top X) + \xi s_1(u_1^\top + u_2^\top X)$$
$$= D - CX + \eta v_1(r_1^\top + r_2^\top X).$$

Together with the fact that

$$(D - CX)v_1 = (\Gamma - qe^\top)\Gamma^{-1}q - qq^\top\Delta^{-1}e = 0,$$

and

$$(r_1^\top + r_2^\top X)v_1 = e^\top\Gamma^{-1}q + q^\top\Delta^{-1}e = 1,$$

we obtain

$$(\overline{D} - \overline{C}X)v_1 = (\overline{D} - \overline{C}X)v_1 = \eta v_1.$$

Then, Lemma 3.1 and Theorem 2.5 imply that $\sigma(\overline{D} - \overline{C}X) = \{\eta, \lambda_2, \ldots, \lambda_n\}$.

Since $\overline{M}$ is a nonsingular M-matrix and $\overline{X}$ is the minimal nonnegative solution of (3.11a), Theorem 2.3 and Theorem 2.4 imply that $\sigma(\overline{D} - \overline{C}\,\overline{X}) \subset \mathbb{C}_+$. With this in mind, we have

$$\sigma(\overline{D} - \overline{C}\,\overline{X}) = \sigma(\overline{D} - \overline{C}X). \tag{3.29}$$

Note that

$$\overline{H}\begin{bmatrix} I_n \\ \overline{X} \end{bmatrix} = \begin{bmatrix} I_n \\ \overline{X} \end{bmatrix}(\overline{D} - \overline{C}\,\overline{X}). \tag{3.30}$$

By (3.29) and (3.30), it is true that

$$\text{span}\begin{bmatrix} I_n \\ X \end{bmatrix} = \text{span}\begin{bmatrix} I_n \\ \overline{X} \end{bmatrix}.$$

Then, there exists a nonsingular matrix $S \in \mathbb{R}^{n \times n}$ such that

$$\begin{bmatrix} I_n \\ X \end{bmatrix} = \begin{bmatrix} I_n \\ \overline{X} \end{bmatrix}S.$$

It is clear that this nonsingular matrix S is an identity matrix. So, we conclude that $X = \overline{X}$.
□

From Theorem 3.7 and Theorem 3.8, we know that $\overline{M}$ is a nonsingular M-matrix. Then, the SDA algorithm is guaranteed to converge. Similar to the discussion given in the single shifted algorithm, we have

$$\limsup_{k\to\infty} \sqrt[2^k]{\|H_k - X\|} \le \rho(C_\gamma(\overline{D} - \overline{C}X))\rho(C_\gamma(\overline{A} - \overline{B}Y)) < 1, \tag{3.31}$$

since $\rho(C_\gamma(\overline{D} - \overline{C}X)) < 1$ and $\rho(C_\gamma(\overline{A} - \overline{B}Y)) < 1$. This also implies that for any $(\eta, \xi) \in \Omega$, the SDA algorithm with double shifts converges faster than the SDA algorithm with no shift and is quadratically convergent.

4. Advantages of the Shifting Technique Applied to SI. In [15], Lu shows that the minimal nonnegative solution X of (1.1) must be of the form:

$$X = T \circ (mn^\top) = (mn^\top) \circ T.$$

Here, the symbol $\circ$ is the Hadamard product, $T = (t_{ij}) = \left(\frac{1}{\delta_i + \gamma_j}\right)$, and (m, n) is satisfying the vector equation:

$$m = m \circ (Pn) + e, \tag{4.1a}$$
$$n = n \circ (Qm) + e, \tag{4.1b}$$

with

$$P = (P_{ij}) = \left(\frac{q_j}{\delta_i + \gamma_j}\right), \quad Q = (Q_{ij}) = \left(\frac{q_j}{\delta_j + \gamma_i}\right). \tag{4.2}$$

The SI method for finding the minimal nonnegative solution (m, n) is then given by

$$m^{(k+1)} = m^{(k)} \circ (Pn^{(k)}) + e, \tag{4.3a}$$
$$n^{(k+1)} = n^{(k)} \circ (Qm^{(k)}) + e. \tag{4.3b}$$

Our aim in this section is to discuss how the shifted approaches can speed up the SI method. Theoretical discussion is also given to analyze the convergence of the SI method with shift. We then rewrite the coefficient matrices (3.12) as

$$\overline{D} = \overline{D}(\eta, \xi) = \Gamma - Q_1(\eta)E_1(\xi)^\top, \tag{4.4a}$$
$$\overline{C} = \overline{C}(\eta, \xi) = Q_1(\eta)Q_2(\xi)^\top, \tag{4.4b}$$
$$\overline{B} = \overline{B}(\eta, \xi) = E_2(\eta)E_1(\xi)^\top, \tag{4.4c}$$
$$\overline{A} = \overline{A}(\eta, \xi) = \Delta - E_2(\eta)Q_2(\xi)^\top, \tag{4.4d}$$

with

$$Q_1 = Q_1(\eta) = \begin{bmatrix} (I_n - \eta\Gamma^{-1})q & q \end{bmatrix}, \quad Q_2 = Q_2(\xi) = \begin{bmatrix} q & \xi\Delta^{-1}q \end{bmatrix},$$
$$E_1 = E_1(\xi) = \begin{bmatrix} e & -\xi\Gamma^{-1}e \end{bmatrix}, \quad E_2 = E_2(\eta) = \begin{bmatrix} (I_n + \eta\Delta^{-1})e & e \end{bmatrix},$$

and relax the boundary conditions (η, ξ) so that $(\eta, \xi) \in \overline{\Omega}$. Here, $\overline{\Omega}$ is the closure of the set Ω defined in (3.16). Substituting (4.4) into (3.11a), we have

$$Z\Gamma + \Delta Z = (ZQ_1 + E_2)(Q_2^\top Z + E_1^\top). \tag{4.5}$$

This implies that the minimal nonnegative solution Z of (3.11a) can be written as

$$Z = T \circ (MN^\top), \tag{4.6}$$

with $M = ZQ_1 + E_2 \in \mathbb{R}^{n \times 2}$, $N^\top = Q_2^\top Z + E_1^\top \in \mathbb{R}^{2 \times n}$.

Akin to the iteration given in (4.3), the iteration sequence $\{M_k, N_k\}$ corresponding to (4.6) can be written as

$$M_{k+1} = (T \circ (M_k N_k^\top))Q_1 + E_2, \tag{4.7a}$$

$$N_{k+1} = (T \circ (N_k M_k^\top))Q_2 + E_1, \tag{4.7b}$$

with the initial value

$$M_0 = 0, N_0 = 0. \tag{4.7c}$$

Let $Z_k(\eta, \xi) = Z_k = T \circ (M_k N_k^\top)$, for all k. Corresponding to (4.5), we then have the classical fixed-point iteration,

$$Z_{k+1} \equiv T \circ \left((Z_k Q_1 + E_2)(Q_2^\top Z_k + E_1^\top)\right). \tag{4.8}$$

Our next theorem is to show that the sequence $\{Z_k\}$ does indeed converge and converge to the minimal nonnegative solution X of (1.1).

THEOREM 4.1. *Assume that*

$$\overline{R}(X^*) = X^* \overline{C} X^* - X^* \overline{D} - \overline{A} X^* + \overline{B} \leq 0, \tag{4.9}$$

for some nonnegative matrix X^*. Then for the fixed-point iteration (4.8) with initial value $Z_0 = 0$, we have

$$Z_0 < Z_1 < \cdots < Z_k < X^*, \text{ for any } k \geq 1. \tag{4.10}$$

Moreover, $\lim\limits_{k \to \infty} Z_k(\eta, \xi) = X$ *for any* $(\eta, \xi) \in \Omega$.

Proof. By (4.4), $Q_1 E_1^\top \geq 0$, $Q_1 Q_2^\top = \overline{C} \geq 0$, $E_2 E_1^\top = \overline{B} \geq 0$, and $E_2 Q_2 \geq 0$. It follows that (4.10) holds by induction. Since the sequence $\{Z_k\}$ is monotonically increasing and bounded above, we have $\lim\limits_{k \to \infty} Z_k = Z^*$, for some Z^*. Hence $\overline{R}(Z^*) = 0$. On the other hand, since $Z^* \leq X^*$ for any nonnegative matrix X^*, we have $Z^* = X$. $\square$

The convergence property, shown in Theorem 4.1, is of fundamental importance in our subsequence discussion and can induce the possibility of analyzing a number of convergent behaviors in the SI method with shift. Note that since M_k and N_k are matrices in $\mathbb{R}^{n \times 2}$, we can define

$$M_k = \begin{bmatrix} m_1^{(k)} & m_2^{(k)} \end{bmatrix}, \ N_k = \begin{bmatrix} n_1^{(k)} & n_2^{(k)} \end{bmatrix}, \tag{4.11}$$

where $m_1^{(k)}, m_2^{(k)}, n_1^{(k)}$ and $n_2^{(k)}$ are n-dimension column vectors. It follows that we have the equivalent iteration for Z_k, that is,

$$Z_k = T \circ \left(m_1^{(k)}(n_1^{(k)})^\top + m_2^{(k)}(n_2^{(k)})^\top\right). \tag{4.12}$$

Then, we obtain the new algorithm of the SI with shift, given by

$$m_1^{(k+1)} = Z_k(I_n - \eta\Gamma^{-1})q + (I_n + \eta\Gamma^{-1})e, \tag{4.13a}$$

$$m_2^{(k+1)} = Z_k q + e, \tag{4.13b}$$

$$n_1^{(k+1)} = Z_k^\top q + e, \tag{4.13c}$$

$$n_2^{(k+1)} = -\xi(\Gamma^{-1}e - Z_k^\top\Delta^{-1}q). \tag{4.13d}$$

with the initial value

$$m_1^{(0)} = 0, \ m_2^{(0)} = 0, \tag{4.14a}$$

$$n_1^{(0)} = 0, \ n_2^{(0)} = 0. \tag{4.14b}$$

It is true that this SI iteration with shift is still a method with a cost of $O(n^2)$ ops but requires more calculations than the original SI method. However, in order to have a method with a better behavior, adding some complexity is sometimes a necessary sacrifice. Actually, we can simplify our computation by consider the following iteration,

$$m_2^{(k+1)} = \begin{bmatrix} Z_k & I_n \end{bmatrix} \begin{bmatrix} q \\ e \end{bmatrix},$$

$$m_1^{(k+1)} = m_2^{(k+1)} + \eta \begin{bmatrix} Z_k & I_n \end{bmatrix} \begin{bmatrix} -\Gamma^{-1}q \\ \Delta^{-1}e \end{bmatrix},$$

$$n_1^{(k+1)} = \begin{bmatrix} Z_k^\top & I_n \end{bmatrix} \begin{bmatrix} q \\ e \end{bmatrix},$$

$$n_2^{(k+1)} = -\xi \begin{bmatrix} Z_k^\top & I_n \end{bmatrix} \begin{bmatrix} -\Delta^{-1}q \\ \Gamma^{-1}e \end{bmatrix}.$$

In next theorem, we discuss the convergent property of the sequence $(m_1^{(k)}, m_2^{(k)}, n_1^{(k)}, n_2^{(k)})$ and the convergent speed of the sequence Z_k.

THEOREM 4.2. *Given $(\alpha, c) = (0, 1)$, the sequence $(m_1^{(k)}, m_2^{(k)}, n_1^{(k)}, -n_2^{(k)})$ with initial values(4.14) is strictly monotonically increasing and satisfies the following two conditions:*

a. $e \le m_1^{(k)} \le m$, $e \le m_2^{(k)} \le m$, $e \le n_1^{(k)} \le n$, $0 \le n_2^{(k)} \le -\xi\Gamma^{-1}e$.

b. $\lim_{k\to\infty} m_1^{(k)} = \lim_{k\to\infty} m_2^{(k)} = m$, $\lim_{k\to\infty} n_1^{(k)} = n$, $\lim_{k\to\infty} n_2^{(k)} = 0$,

where m and n are defined on (4.1). In fact, in the critical case, we have $m = n$ and $X = X^\top$.

Proof. From Theorem 4.1, we know that $Z_0 < Z_1 < \ldots < Z_k < X$ and $\lim_{k\to\infty} Z_k = X$. Substituting these two facts to (4.13), we immediately have

$$
\begin{array}{ccccccccc}
e & \le & m_2^{(1)} & < & m_2^{(2)} & < & \ldots & < & m_2^{(k)} & \le & Xq + e = m, \\
e & \le & m_1^{(1)} & < & m_1^{(2)} & < & \ldots & < & m_1^{(k)} & \le & Xq + e = m, \\
e & \le & n_1^{(1)} & < & n_1^{(2)} & < & \ldots & < & n_1^{(k)} & \le & X^\top q + e = n, \\
-\xi\Gamma^{-1}e & \ge & n_2^{(1)} & > & n_2^{(2)} & > & \ldots & > & n_2^{(k)} & \ge & -\xi(\Gamma^{-1}e - X^\top\Delta^{-1}q) = 0.
\end{array}
$$

Note that the order of the sequence$\{m_1^{(k)}\}$ comes from the fact that $Z_k q - \eta Z_k\Gamma^{-1}q > 0$ and the last equality of the sequence $\{n_2^{(k)}\}$ comes from Theorem 2.6.

From (4.12), part (b) holds, since $\lim_{k\to\infty} Z_k = X$. □

When we studied the shifted procedures, our main purpose is to speed up the convergence. In what follows we discuss the relations of $Z_k(\eta, \xi)$ with respect to different η and ξ values and show that the SI with shift converges linear, instead of sublinear.

THEOREM 4.3. *Given $(\alpha, c) = (0, 1)$, the sequence $\{Z_k\}$ has the following two properties:*

a. $Z_k(0,0) \leq Z_k(\eta,0) \leq Z_k(\eta,\xi)$, *for each k and $(\eta,\xi) \in \Omega$.*

b. *The sequence $\{Z_k(\eta,\xi)\}$ converges linearly to the minimal nonnegative solution X of (1.1) for all $(\eta,\xi) \in \Omega$.*

Proof. From (4.6), we have

$$
\begin{aligned}
Z &= T \circ \left(ZQ_1(\eta)Q_2(\xi)^\top Z + E_2(\eta)Q_2(\xi)^\top Z + ZQ_1(\eta)E_1(\xi)^\top + E_2(\eta)E_1(\xi)^\top \right) \\
&= T \circ \left((ZQ_1(0)Q_2(0)^\top Z + E_2(0)Q_2(0)^\top Z + ZQ_1(0)E_1(0)^\top + E_2(0)E_1(0)^\top \right) \\
&\quad + \eta T \circ \left((-Z\Gamma^{-1}q + \Delta^{-1}e)(q^\top Z + e^\top) \right) - \xi T \circ \left((Zq + e)(-q^\top \Delta^{-1} Z + e^\top \Gamma^{-1}) \right).
\end{aligned}
$$

Subsequently, it follows from mathematical induction that part (a) holds.

For the proof of part (b), we need to use three well-known results discussed in [9]. First, for the iteration (4.8) and $Z_0(\eta,\xi) = 0$, we have

$$
\begin{aligned}
&\limsup_{k \to \infty} \sqrt[k]{\|Z_k(\eta,\xi) - X\|} \\
&= \rho \left((I \otimes \Delta + \Gamma \otimes I)^{-1} \left[I \otimes (E_2 E_1^\top + X\overline{C}) + (Q_1 E_1^\top + \overline{C}X) \otimes I \right] \right), \qquad (4.15)
\end{aligned}
$$

where $\otimes$ denotes the Kronecker product (see [9, Theorem 3.2]). Second, let $M_X = I \otimes (\overline{A} - X\overline{C}) + (\overline{D} - \overline{C}X)^\top \otimes I$. Then, M_X is a Z-matrix since both $\overline{A} - X\overline{C}$ and $\overline{D} - \overline{C}X$ are Z-matrices. (see [9, Remark 1.1]). Also, M_X is a nonsingular matrix since any eigenvalue of M_X is the sum of an eigenvalue of $\overline{A} - X\overline{C}$ and $\overline{D} - \overline{C}X$. This implies that M_X is a nonsingular M-matrix. Third, if M_X is a nonsingular M-matrix, then

$$
\rho \left((I \otimes \Delta + \Gamma \otimes I)^{-1} \left[I \otimes (E_2 E_1^\top + X\overline{C}) + (Q_1 E_1^\top + \overline{C}X) \otimes I \right] \right) < 1, \qquad (4.16)
$$

that is, $\limsup_{k \to \infty} \sqrt[k]{\|Z_k(\eta,\xi) - X\|} < 1$. (see [9, Theorem 3.3])

$\square$

5. Numerical Implementation and Comparisons. To illustrate the consequence of the previous sections, numerical experiments, consisting of SDA and SI methods after the shifting technique, are presented to demonstrate our conclusion. All computations are performed in MATLAB/version 2010b on a iMac with an 2.8GHZ Intel Core i5 processor and 16GB main memory, using IEEE double-precision.

In the next implementations, the relative error for the SDA is defined by

$$
\text{Err}_{SDA} = \max \left\{ \frac{\|G_k - G_{k-1}\|_\infty}{\|G_k\|_\infty}, \frac{\|H_k - H_{k-1}\|_\infty}{\|H_k\|_\infty} \right\},
$$

the relative error for the SI with no shift is defined by

$$
\text{Err}_{SI} = \max \left\{ \frac{\|m^{(k)} - m^{(k-1)}\|_\infty}{\|m^{(k)}\|_\infty}, \frac{\|n^{(k)} - n^{(k-1)}\|_\infty}{\|n^{(k)}\|_\infty} \right\},
$$

the relative error for the SI with the shifting procedure is defined by

$$
\text{Err}_{SIS} = \max \left\{ \frac{\|M_k - M_{k-1}\|_\infty}{\|M_k\|_\infty}, \frac{\|N_k - N_{k-1}\|_\infty}{\|N_k\|_\infty} \right\},
$$

and the relative normalized residual is defined by

$$
\text{Res} = \frac{\|X_k\Gamma + \Delta X_k - (X_k q + e)(q^\top X_k + e^\top)\|_\infty}{\|X_k\|_\infty\|\Gamma\|_\infty + \|X_k\|_\infty\|\Delta\|_\infty + (\|X_k\|_\infty\|q\|_\infty + \|e\|_\infty)(\|q^\top\|_\infty\|X_k\|_\infty + \|e^\top\|_\infty)},
$$

where $X_k = G_k$ for the SDA algorithm, $X_k = T \circ (m^{(k)}n^{(k)^\top})$ for the SI algorithm with no shift and $X_k = T \circ (M_k N_k^\top)$ for the SI algorithm with shift. All iteration methods are terminated whenever the relative errors or the relative normalized residual residuals are less than $n^2\epsilon$, where $\epsilon = 2^{-52} \cong 2.22 \cdot 10^{-16}$ be the machine zero.

EXAMPLE 5.1. *In this example, we compare the methods for finding the minimal nonnegative solution of (1.1) by using the shifting technique. We explain the efficiency of the SDA and SI applied to the shifted equations (3.3) and (3.12), respectively. We consider (1.1) with $(\alpha, c) = (0, 1)$. As suggested in [9, 14], the constants c_i and ω_i are the nodes and weights, which are obtained by dividing the interval $[0, 1]$ into $n/4$ subinterval of equal length and applying to each subinterval the 4-node Gauss–Legendre quadrature.*

In table 5, we report a comparison of residuals and the number of iterations for the SDA with no shift, the SDA with a single shift, the SDA with double shifts , the SI with no shift, the SI with a single shift, and the SI with double shifts and with size $n = 32, 64, 128$, and 256. From table 5, we have the following two conclusions.

First, in the critical case $(\alpha, c) = (0, 1)$, it is known that the SDA algorithm converges linearly. After applied to the shifted equation, the SDA algorithm converges quadratically. As shown in Table 5, the number of steps required in the SDA algorithm with a single shift or double shifts are around half of those of the SDA algorithm with no shift. Also, the computed solution of the shifted equations is more accurate than the one obtained with no shift. The numerical phenomena are in accordance with the theoretical discussion given in [7].

Second, we randomly choose η and ξ from the set Ω. Indeed, in Table 5, we have $(\eta, \xi) = (\frac{1}{2\omega_1}, 0)$ for the single-shift problems and $(\eta, \xi) = (\frac{1}{2\omega_1}, \frac{-1}{2\omega_1})$ for the double-shift problems. We see that even with 10000 steps, the solution obtained from the nonshifted problems can only have accuracy up to 10^{-8}. On the other hand, the solution for the shifted problems can have the accuracy better than 10^{-10} and a dramatical decrease in the number of steps required in the computation. Also, the iteration counts listed in Table 5 are in accord with Theorem 4.3.

n	SDA(no shift)	SDA(single shift)	SDA(double shifts)
32	9.7e-14(27)	4.5e-15(11)	7.4e-15(11)
64	4.2e-13(27)	1.6e-14(12)	1.9e-14(12)
128	1.7e-12(27)	4.2e-14(13)	6.1e-14(13)
256	6.8e-12(27)	1.2e-13(14)	1.4e-13(14)
n	SI(no shift)	SI(single shift)	SI(double shifts)
32	* ($>$10000)	2.4e-13(164)	2.9e-13(40)
64	*($>$10000)	1.0-12(154)	1.3e-12(38)
128	*($>$10000)	4.0-12(145)	5.4e-12(36)
256	*($>$10000)	1.6-11(136)	2.2e-11(34)

TABLE 5.1
Comparison of the residuals (and in parentheses the number of iterations) of the SDA and SI techniques.

6. Conclusion. The challenging issues of applying the SDA algorithm to the shifted NARE problems are to develop a well-defined sequence, to guarantee the convergence of the sequence, and to associate the solutions of the shifted problems with the original one. All these issues related to the structed NARE (1.1) have been studied in our work. Numerical experiments show the improvement of the speed and accuracy while applying the SDA algorithm to the shifted problems. Note that the bottleneck for applying this algorithm is to compute the inverses of $(I_n - H_k G_k)$ and $(I_n - G_k H_k)$, which apparently have an $O(n^3)$ complexity. Compare with

the Newton method, which has been shown to have $O(n^2)$ complexity [4], an interesting problem worthy of further investigation is to reduce the computational cost by taking the specific structure of (1.1) into account.

On the other hand, while applying the SI algorithm to the critical case, its convergence is very slow and has almost stopped. Through the shifting technology, a new iteration method has been introduced and preserve the linear convergence. Numerical experiments show that while considering the SI algorithm, the convergence with double shifts is much faster than the convergence with a single shift or no shift. We believe the results we obtain are new in the field and could provide considerable insight into the NARE problems.

REFERENCES

[1] Z.-Z. BAI, Y.-H. GAO, AND L.-Z. LU, *Fast iterative schemes for nonsymmetric algebraic riccati equations arising from transport theory*, SIAM J. Sci. Comput., 30 (2008), pp. 804–818.

[2] L. BAO, Y. LIN, AND Y. WEI, *A modified simple iterative method for nonsymmetric algebraic Riccati equations arising in transport theory*, Appl. Math. Comput., 181 (2006), pp. 1499–1504.

[3] A. BERMAN AND R. J. PLEMMONS, *Nonnegative matrices in the mathematical sciences*, vol. 9 of Classics in Applied Mathematics, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1994. Revised reprint of the 1979 original.

[4] D. A. BINI, B. IANNAZZO, AND F. POLONI, *A fast Newton's method for a nonsymmetric algebraic Riccati equation*, SIAM J. Matrix Anal. Appl., 30 (2008), pp. 276–290.

[5] C.-Y. CHIANG, E. K.-W. CHU, C.-H. GUO, T.-M. HUANG, W.-W. LIN, AND S.-F. XU, *Convergence analysis of the doubling algorithm for several nonlinear matrix equations in the critical case*, SIAM J. Matrix Anal. Appl., 31 (2009), pp. 227–247.

[6] C.-H. GUO, *Nonsymmetric algebraic Riccati equations and Wiener-Hopf factorization for M-matrices*, SIAM J. Matrix Anal. Appl., 23 (2001), pp. 225–242 (electronic).

[7] C.-H. GUO AND N. J. HIGHAM, *Iterative solution of a nonsymmetric algebraic Riccati equation*, SIAM J. Matrix Anal. Appl., 29 (2007), pp. 396–412.

[8] C.-H. GUO, B. IANNAZZO, AND B. MEINI, *On the doubling algorithm for a (shifted) nonsymmetric algebraic Riccati equation*, SIAM J. Matrix Anal. Appl., 29 (2007), pp. 1083–1100.

[9] C.-H. GUO AND A. J. LAUB, *On the iterative solution of a class of nonsymmetric algebraic Riccati equations*, SIAM J. Matrix Anal. Appl., 22 (2000), pp. 376–391 (electronic).

[10] C.-H. GUO AND W.-W. LIN, *Convergence rates of some iterative methods for nonsymmetric algebraic Riccati equations arising in transport theory*, Linear Algebra Appl., 432 (2010), pp. 283–291.

[11] X.-X. GUO, W.-W. LIN, AND S.-F. XU, *A structure-preserving doubling algorithm for nonsymmetric algebraic Riccati equation*, Numer. Math., 103 (2006), pp. 393–412.

[12] R. A. HORN AND C. R. JOHNSON, *Topics in matrix analysis*, Cambridge University Press, Cambridge, 1991.

[13] J. JUANG AND W.-W. LIN, *Nonsymmetric algebraic Riccati equations and Hamiltonian-like matrices*, SIAM J. Matrix Anal. Appl., 20 (1999), pp. 228–243 (electronic).

[14] L.-Z. LU, *Newton iterations for a non-symmetric algebraic Riccati equation*, Numer. Linear Algebra Appl., 12 (2005), pp. 191–200.

[15] ———, *Solution form and simple iteration of a nonsymmetric algebraic Riccati equation arising in transport theory*, SIAM J. Matrix Anal. Appl., 26 (2005), pp. 679–685 (electronic).

[16] R. S. VARGA, *Matrix iterative analysis*, vol. 27 of Springer Series in Computational Mathematics, Springer-Verlag, Berlin, expanded ed., 2000.